A New Framework for Fusing Stereo Images with Volumetric Medical Images

Fabienne Betting*, Jacques Feldmar*, Nicholas Ayache*, Frédéric Devernay+

INRIA SOPHIA, *Projet Epidaure,+Projet Robotvis
2004 route des Lucioles, B.P. 93
06902 Sophia Antipolis Cedex, France.
Email : Jacques.Feldmar@sophia.inria.fr

Abstract. Some medical interventions require knowing the correspondence between an MRI/CT image and the actual position of the patient. Examples are in neurosurgery or radiotherapy, but also in video surgery (laparoscopy). Recently, computer vision techniques have been proposed to find this correspondence without any artificial markers. Following the pioneering work of [GLPI+94], [CZH+94], [CDT+92], [SHK94] and [STAL94], we propose in this paper an alternative approach.
We propose to trade the laser range finder for two cameras. Hence, we get dense reconstruction of the patient's surface and this allows us to compute the normals to the surface. We present a new method for rigid registration when surfaces are described by points and normals. It does not depend on the initial positions of the surfaces, deals with occlusion in a strict way and takes advantage of the normal information.
Results are presented on real images.

1 Introduction

Medical images are commonly used to help establish a correct diagnosis. As they contain spatial information, both anatomical and functional, they can also be used to planify therapy, and even in some cases to control the therapy.

A recent overview of these fields of research can be found in [Aya93] and in [TLBM94], a spectacular use of planification and control of therapy using medical images and robots can be found in [Tay93, CDT+92, LSB91, LC90] for surgery and [STAL94] for radiotherapy.

More recently, the possibility of helping the surgeon to control the therapy with a projection of some pre-operative images directly onto the patient during surgery, was presented in [GLPI+94, CZH+94, SHK94].

The idea is to present on the patient some anatomical or pathological structures segmented onto pre-operative images, which are difficult to observe during surgery because they are still hidden, or simply because they are not immediately visible under normal lighting conditions (a tumor may be much more visible on a particular MRI image than with direct observation). This process is called *enhanced visualization* or *augmented reality.*

To project an image on a patient, a future solution will probably be the use of semi-transparent glasses or screen, allowing both the direct observation and the

chosen image projection. A somewhat simpler solution consists in acquiring a video image of the patient from a point of view similar to the surgeon's one, and to fuse the chosen pre-operative image with this video image. Such a solution was described in [GLPI+94].

The difficult task is then to register, if possible in real time, the video image of the patient (intra-operative image) with the pre-operative image. This problem could be solved with artificial marker visible in both images, but this produces unacceptable constraints (e.g. pre-operative images must be taken the same day as the intervention, stereotactic frames are very painful to bear and can prevent a free access to the surgeon, etc...).

The registration problem is solved without any artificial marker by [GLPI+94] and [SHK94] with an intermediate laser range finder, which provides a 3D description of the patient's surface. This surface is then matched against the surface of the segmented corresponding surface in the volumetric medical image. As the laser range finder is calibrated with respect to the camera, the medical image can be fused with the video image.

In this paper we propose an alternative solution, where we would change the laser range finder for two cameras, and build the patient's surface with passive stereovision [DF94, Kan93]. Note that [CZH+94] also uses a stereo system.

The advantages of such an approach are the following: first, we use passive vision, instead of a laser beam, which can be annoying in the surgery room. Second, having stereo cameras, the final enhanced visualization can be done in stereovision, providing a much more vivid representation of the patient's anatomy. Third, we get a dense reconstruction of the patient's surface and this allows us to compute the normals to the surface. Thanks to these normals, we may introduce a completely new method for rigid registration when surfaces are described by points and normals as it can often be the case in medical imaging. Fourth, thanks to the density of the surface descriptions, we believe that we can get a more accurate registration than the techniques using sparse data.

The paper is organized as follows: in section 2 we briefly describe the geometry of the registration problem. In section 3, we introduce our new algorithm to find an estimate of the rigid displacement based on bitangent segments and we analyze the complexity of this algorithm. In section 4, we present a new distance minimization algorithm, which is an extension of the iterative closest point algorithm, dealing with occlusion in a strict manner and taking advantage of the normal information. Experimental results are presented in section 5. Finally, future work is presented in conclusion.

2 Geometry of the registration problem

Before the intervention, an MRI/CT image of the patient's head is acquired. This image contains the skin and the brain of the patient and also a possible tumor. The goal is to find, during the intervention, the projective transformation which maps this MRI/CT image on a video image of the patient's head.

Following the approach presented in [GLPI+94, CZH+94, SHK94], we also split the problem in two stages: reconstruction and rigid registration. But the

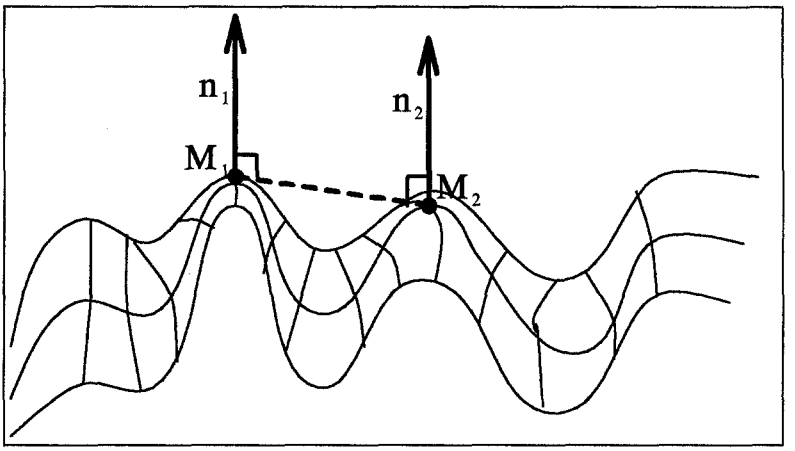

Fig. 1. A surface and two bitangent points M_1 and M_2. Let $\mathbf{n}_1$ and $\mathbf{n}_2$ be the normals at these points. M_1 and M_2 are bitangent if the plane defined by $(M_1, \mathbf{n}_1)$ and the plane defined by $(M_2, \mathbf{n}_2)$ are the same. Another definition is that $\mathbf{n}_1$ and $\mathbf{n}_2$ are identical and that the line M_1M_2 is orthogonal to these two vectors.

way we perform the surface reconstruction is different. We use passive stereo as described in [DF94]. The result is a dense description of the patient's surface by points and normals. The coordinates of these points and normals are expressed in the camera frame. Because the transformation which maps the reconstructed image to the camera image is known, the problem is to find the transformation between the MRI/CT image and the reconstructed surface.

In order to find this transformation, we extract in the MRI/CT image the patient's surface and we get a description by points and normals. The rigid registration problem is the following:
Given two surfaces described by points and normals, find the rigid displacement that best superposes these two surfaces.

As mentioned in [GLPI+94], this algorithm must not depend on the initial relative positions of the surfaces, it must be accurate and robust.

3 Finding an initial estimate of the rigid displacement

The basic idea is to compute independently on each surface the set of pairs of points sharing the same tangent plane (see figure 1). We call such pairs **bitangent points**. They correspond to semi-differential invariants [GMPO92]. The technique for computing these pairs is described in [FAF94]. We simply note here that the algorithm is quasi-linear in the number of points describing the surface, and, because it involves only derivatives of order 1, the bitangent points calculation is quite stable.

In the ideal case, because the distance between the two bitangent points is invariant under rigid displacement, the following algorithm would be very effi-

cient to rigidly superpose a surface S_1 on a surface S_2:
(1) choose a pair P_1 of bitangent points on S_1. Let $d(P_1)$ be the distance between the two points.
(2) Compute the set $SameDistance(P_1)$ of pairs of bitangent points on S_2 such that the distance between the two bitangent points is equal to $d(P_1)$.
(3) For each pair P_2 in $SameDistance(P_1)$, compute the two possible rigid displacements corresponding to the superposition of the two pairs P_1 and P_2 and of their normals. Stop when the rigid displacement which superposes S_1 on S_2 is found.

In practice, corresponding pairs of bitangent points cannot be exactly superposed because of the point discretization error and because of the error in the computation of the normal. Moreover, only a part of the reconstructed surface S_1 may be superposed on the patient's surface extracted from the MRI image S_2. So, the actual algorithm is slightly more complex, but is basically as stated. A detailed description and an analysis of the complexity can be found in [FAF94]. We simply note that the complexity of the algorithm is quasi-linear in the number of points on the surfaces and that the risk of stopping the algorithm with a wrong initial estimate decreases extremely quickly with the number of points (when the two surfaces actually show some overlapping regions up to a rigid displacement).

4 The new distance minimization algorithm

4.1 The Iterative Closest Point algorithm

Using the pairs of bitangent points as described in the previous section, we get an estimate $(\mathbf{R}_0, \mathbf{t}_0)$ of the rigid displacement to superpose S_1 on S_2. In order to find an accurate rigid displacement we have developed an extension of an algorithm called "the Iterative Closest Point algorithm" which was introduced by several researchers ([BM92], [Zha94], [CM92], [CLSB92]). We sketch the original ICP algorithm, which searches for the rigid displacement $(\mathbf{R}, \mathbf{t})$ which minimizes the energy

$$E(\mathbf{R}, \mathbf{t}) = \sum_{M_i \in S_1} \|\mathbf{R}M_i + \mathbf{t} - ClosestPoint(\mathbf{R}M_i + \mathbf{t})\|^2,$$

where $ClosestPoint$ is the function which associates to a space point its closest point on S_2.

The algorithm consists of two iterated steps, each iteration i computing a new estimation $(\mathbf{R}_i, \mathbf{t}_i)$ of the rigid displacement.

1. The first step builds a set $Match_i$ of pairs of points: for each point M on S_1, a pair (M, N) is added to $Match_i$, where N is the closest point on S_2 to the point $\mathbf{R}_{i-1}M + \mathbf{t}_{i-1}$.
2. The second step is the least squares evaluation of the rigid displacement $(\mathbf{R}_i, \mathbf{t}_i)$ to superpose the pairs of $Match_i$.

The termination criterion depends on the approach used: the algorithm stops either when a) the distance between the two surfaces is below a fixed threshold, b) the variation of the distance between the two surfaces at two successive iterations is below a fixed threshold or c) a maximum number of iterations is reached.

This ICP algorithm is efficient and finds the correct solution when the initial estimate $(\mathbf{R}_0, \mathbf{t}_0)$ of the rigid displacement is "not too bad" and when each point on S_1 has a correspondent on S_2. But in practice, this is often not the case. For example in our application, as explained in the previous section, the reconstructed surface usually only describes partially the patient's surface and often includes a description of the patient's environment. The next two subsections explain how we deal with these two problems.

4.2 Working with incomplete surfaces

In step 1 of the iterative algorithm, we map each point of S_1 to a "closest point" on S_2. But when the two surfaces are partially reconstructed, some points on S_1 do not have any homologous point on S_2. Thus, given a point M on S_1, $(\mathbf{R}_{i-1}, \mathbf{t}_{i-1})$, and $ClosestPoint(\mathbf{R}_{i-1}M + \mathbf{t}_{i-1})$, we have to decide whether $(M,\ ClosestPoint(\ \mathbf{R}_{i-1}M\ + \mathbf{t}_{i-1}\ )\ )$ is a plausible match. This is very important because, if we accept incorrect matches, the found rigid displacement will be biased (and therefore inaccurate), and if we reject correct matches, the algorithm may not converge towards the best solution.

As proposed in [Aya91], we make use of the extended Kalman filter (EKF). This allows us to associate to the six parameters of $(\mathbf{R}_i, \mathbf{t}_i)$ a covariance matrix $\mathbf{S}_i$ and to compute a generalized Mahalanobis distance δ for each pair of matched points (M, N). This generalized Mahalanobis distance, under some assumptions on the noise distributions and some first-order approximations, is a random variable with a χ^2 probability distribution. By consulting a table of values of the χ^2 distribution, it is easy to determine a confidence level ϵ for δ corresponding to, for example a 95% probability of having the distance δ less than ϵ. In this case, we can consider the match (M, N) as *likely* or *plausible* when the inequality $\delta < \epsilon$ is verified and consider any others as *unlikely* or *unplausible.*

This distinction between plausible and unplausible matches implies a change in the second step of the iterative algorithm. Given $Match_i$, instead of computing the rigid displacement $(\mathbf{R}, \mathbf{t})$ which minimizes the least squares criterion

$$\sum_{(M,N)\ \in\ Match_i} \|\mathbf{R}_i M + \mathbf{t}_i - N\|^2,$$

we recursively estimate the six parameters of $(\mathbf{R}, \mathbf{t})$, and the associated covariance matrix which minimizes the criterion

$$\sum_{(M,N)\ \in\ Match_i\ and\ (M,N)\ is\ plausible} (\mathbf{R}_i M + \mathbf{t}_i - N)^t \mathbf{W}_i^{-1} (\mathbf{R}_i M + \mathbf{t}_i - N),$$

where $\mathbf{W}_i$ is a covariance matrix which allows us, for example, to increase the importance of high curvature points.

More details about the meaning of "plausible or not" and about the EKF can be found in [FAF94].

4.3 Using the normal information

As is commonly encountered with any minimization algorithm, the ICP algorithm may become trapped in a local minimum. To reduce this problem, we propose in this subsection to make use of the normal information and to define a new criterion to minimize. In our formulation, surface points are no longer 3D points: they become 6D points. Coordinates of a point M on the surface S are (x, y, z, n_x, n_y, n_z) where (n_x, n_y, n_z) is the normal to S at M. For two points $M(x, y, z, n_x, n_y, n_z)$ and $N(x', y', z', n'_x, n'_y, n'_z)$ we define the distance:

$$d(M,N) = (\ \alpha_1(x-x')^2 + \alpha_2(y-y')^2 + \alpha_3(z-z')^2 + \\ \alpha_4(n_x-n'_x)^2 + \alpha_5(n_y-n'_y)^2 + \alpha_6(n_z-n'_z)^2)^{1/2}$$

where α_i is the inverse of the difference between the maximal and minimal value of the i^{th} coordinate of points in S_2. Using this definition of distance, the closest point to P on S_2 is a compromise between the 3D distance and the difference in normal orientation[1].

This new definition of the distance between points naturally implies modifications to steps one and two of the ICP algorithm in order to minimize the new energy:

$$E(\mathbf{R},\mathbf{t}) = \sum_{M \in S_1} d(\ (\mathbf{R}M + \mathbf{t}, \mathbf{R}n_1(M)), \\ ClosestPoint\_6D((\mathbf{R}_{i-1}M + \mathbf{t}, \mathbf{R}_{i-1}n_1(M)))^2,$$

where $n_1(M)$ is the normal on S_1 at point M and $ClosestPoint\_6D$ is the new 6D closest point function.

In step one, the closest point now has to be computed in 6D space. We use the kd-tree technique first proposed by Zhang ([Zha94]) for the 3D case. The second step also has to be modified: the criterion which defines the best rigid displacement must use the new 6D distance. Otherwise, it is not possible to prove the convergence of our new ICP algorithm (see [FAF94]). Hence, the rigid displacement $(\mathbf{R}_i, \mathbf{t}_i)$ is now defined as the minimum of the function

$$f(\mathbf{R},\mathbf{t}) = \sum_{(M,N) \in match_i} d(\mathbf{R}M + \mathbf{t}, N)^2,$$

where, the coordinates of the point $\mathbf{R}M + \mathbf{t}$ are $(\mathbf{R}M + \mathbf{t}, \mathbf{R}n_1(M))$.

In practice, we use extended Kalman filters to minimize this new criterion at step 2. Even though it is is non linear, the minimization works very well. Note that this use of extended Kalman filters allows us to compute Mahalanobis distances and to determine if a match is plausible or not as explained in the previous subsection.

[1] Of course, only two parameters are necessary to describe the orientation of the normal (for example the two Euler angles). But we use (n_x, n_y, n_z) because the Euclidean distance better reflects the difference of orientation between the normals (that is not the case with the Euler angles because of the modulo problem) and we can use kd-trees to find the closest point as explained later.

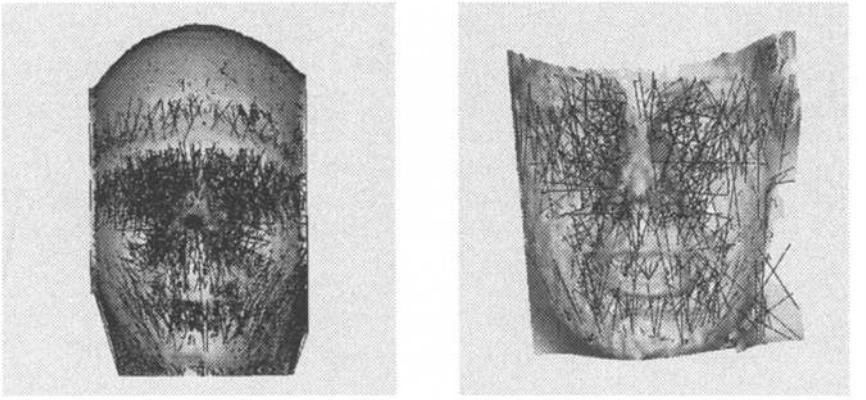

Fig. 2. The bitangent lines computed on the MRI (left) and stereo (right) face surfaces. Note that we selected the lines whose length varies between $2cm$ and $10cm$.

In order to try to demonstrate that the ICP algorithm, using the 6D distance, converges more often to the global minimum than the standard ICP algorithm, we conducted the following experiment. We chose S_1 and S_2 to be the same surface. Hence, the resulting transformation should be the identity. We run both the original and the modified algorithm choosing different initial rigid displacements $(\mathbf{R}_0, \mathbf{t}_0)$, at an increasing distance from the identity. The results are reported in [FAF94] and show that our modified algorithm is in practice much less sensitive to the initial estimate $(\mathbf{R}_0, \mathbf{t}_0)$, and more robust to local minima.

5 Results

We now present an example of application of the framework presented in this section.

First we compute on both the stereo reconstructed surface, and on the MRI surface (figure 2), the pairs of bitangent points. We find 598 pairs on the stereo surface and 5000 pairs on the MRI surface (obviously, the desired density of the bitangent pairs is a parameter of the bitangent extraction algorithm). Hence, there are more pairs on the MRI surface than on the stereo surface which makes the algorithm more efficient. The extraction requires about 30 seconds.[2].

Using these pairs of bitangent points, we estimate the rigid displacement in about 30 seconds. Applying this estimate, 80% of the points on the stereo surface have their closest point at a distance lower than $8mm$. This error has to be compared with the size of the voxel in the MRI image: $4mm \times 4mm \times 2mm$. Moreover, recall that there are points on the stereo surface which do not have a homologous point on the MRI surface.

[2] CPU times are given for a DEC-ALPHA workstation

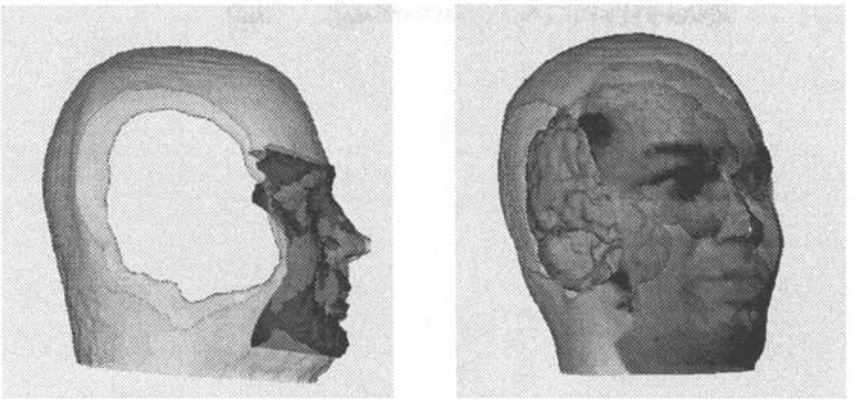

Fig. 3. Left: The result of the rigid registration. The MRI surface is clearer and the stereo surface is darker. The alternation dark/clear shows that the registration is quite accurate. **Right:** The colorized MRI head surface obtained by attaching to each matched point the gray level of its corresponding point on the stereo surface. Thanks to the transparency effect, one can observe the brain...

Using this estimate of the rigid displacement, we run the modified iterative closest point algorithm. The MRI head surface is described by 15000 points and the stereo surface by 10000 points. It takes 20 seconds. Applying this new rigid displacement, 85% of the points on the stereo surface have their closest point at a distance lower than $3mm$. The average distance between matched points is $1.6mm$. The result is presented in figure 3, left.

Because we know the point-to-point correspondences between the MRI head surface and the stereo face surface (this is the result of the registration), and because for each point of the stereo surface we know the grey level from the video image, we can map the video image onto the MRI surface (figure 3, right). The fact that the points on the MRI surface have the right grey levels qualitatively demonstrates that the MRI/stereo matching is correct.

Finally, we projected the brain onto the video image (see figure 4) using the computed projective transformation. In fact, we now have enough geometric and textural parameters to produce a stereo pair of realistic images from a continuous range of viewpoints and provide the surgeon the feeling of seeing inside the patient's head and guide him/her during the intervention.

6 Conclusion and future work

We proposed to use passive stereovision and a new rigid matching algorithm to register pre-operative with intra-operative images. Though we have presented results on real data, a lot of work still has to be done. We would like to compare the techniques described in this paper with respect to the techniques presented by others. For example, we plan to develop procedures to validate rigorously

Fig. 4. The projection of the brain in the two video images using the projective transformation computed as explained in this paper. A stereoscopic display could provide the surgeon with the feeling of seeing inside the patient's head. The two presented images cannot be visually fused in this position, because the baseline between the two optical centers of the video cameras was vertical in this experiment: one can notice that the camera of left image was bellow the camera of the right one. This will be corrected in further experiments.

the accuracy of the methods. We also have to validate our approach on a larger scale, if possible in hospital environment.

We believe that it should be possible to perform real time tracking of the patient and enable the surgeon to move either patient or the 2D sensor. Indeed, for tracking we just need to correct a rigid displacement which is quite close to the right solution. Hence, because the initialization would be good, it should be possible to use the iterative closest point algorithm with just a few reconstructed points without local minimum problem and to get fast convergence. We hope and believe that it should be possible to perform the loop reconstruction/registration/visualization at a $1Hertz$ frequency using still existing fast stereo systems (as [DF94, Kan93]) and efficient graphic hardware (as Kubota or Silicon Graphics).

The passive video system should also allow us to build a system to visualize the surgeon's instruments in the MRI/CT image. Indeed, assume that on each instrument a few points (at least three) are easily identified in the two camera images. We can triangulate these points and find the position of the instruments with respect to the MRI/CT image. We believe that this could be helpful for surgeons, especially for interventions requiring high accuracy.

Acknowledgments: *we wish to thank G. Malandain and M. Brady for practical help and stimulating discussions. Thanks are also due to* **General Electric Medical System Europe (GEMS)** *for providing us with some of the data presented here. Thanks to Dr Michel Royon (Cannes Hospital) and Dr Jean-Jacques Baudet (La Timone Hospital) for fruitful discussions. This work was supported in part by a grant from* **Digital Equipment Corporation** *and by the European Basic Research Action* **VIVA**.

References

[Aya91] N. Ayache. *Artificial Vision for Mobile Robots — Stereo-Vision and Multisensory Perception.* Mit-Press, 1991.

[Aya93] N. Ayache. Analysis of three-dimensional medical images - results and challenges. research report 2050, INRIA, 1993.

[BM92] Paul Besl and Neil McKay. A method for registration of 3−D shapes. *PAMI*, 14(2):239–256, February 1992.

[CDT+92] P. Cinquin, P. Demongeot, J. Troccaz, S. Lavallee, G. Champleboux, L. Brunie, F. Leitner, P. Sautot, B. Mazier, A. Perez M. Djaid, T. Fortin, M. Chenin, and A. Chapel. Igor: Image guided operating robot. methodology, medical applications, results. *ITBM*, 13:373–393, 1992.

[CLSB92] G. Champleboux, S. Lavallée, R. Szeliski, and L. Brunie. From accurate range imaging sensor calibration to accurate model-based 3−D object localization. In *CVPR*, Urbana Champaign, June 1992.

[CM92] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. *Image and Vision Computing*, 10(3):145–155, 1992.

[CZH+94] A.C.F. Colchester, J. Zhao, C. Henri, R.L. Evans, P. Roberts, N. Maitland, D.J. Hawkes, D.L.G. Hill, A.J. Strong, D.G. Thomas, M.J. Gleeson, and T.C.S Cox. Craniotomy simulation and guidance using a stereo video based tracking system (vislan). In *Visualization in Biomedical Computing*, Rochester, Minnesota, October 1994.

[DF94] F. Devernay and O. D. Faugeras. Computing differential properties of 3d shapes from stereoscopic images without 3d models. In *CVPR*, Seattle, USA, June 1994.

[FAF94] J. Feldmar, N. Ayache, and F.Betting. 3d-2d projective registration of free-form curves and surfaces. research report 2434, INRIA, 1994. Available via ftp anonymous on zenon.inria.fr, file /pub/rapports/RR-2434.ps and via WWW ftp://zenon.inria.fr/pub/rapports.

[GLPI+94] W.E.L. Grimson, T. Lozano-Perez, W.M. Wells III, G.J. Ettinger, S.J White, and R. Kikinis. An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization. In *CVPR*, Seattle, USA, June 1994.

[GMPO92] Luc Van Gool, Theo Moons, Eric Pauwels, and André Oosterlinck. Semi-differential invariants. In Joseph L. Mundy and Andrew Zisserman, editors, *Geometric Invariance in Computer Vision.* Mit-Press, 1992.

[Kan93] T. Kanade. very fast 3d sensing hardware. *Internationnal Symposium on Robotic Research*, October 1993.

[LC90] S. Lavallée and P. Cinquin. *Computer assisted medical intervention*, volume 60. Hohne K.H. et al ed, Imaging in Medicine. Springer Verlag, 1990.

[LSB91] Stéphane Lavallée, Richard Szeliski, and Lionel Brunie. Matching 3-d smooth surfaces with their 2-d projections using 3-d distance maps. In *SPIE, Geometric Methods in Computer Vision*, San Diego, Ca, July 1991.

[SHK94] D. Simon, M. Hebert, and T. Kanade. Techniques for fast and accurate intra-surgical registration. In *First international symposium on medical robotics and computer assisted surgery*, Pittsburgh, September 1994.

[STAL94] A. Shweikard, R. Tombropoulos, J. Adler, and J.C. Latombe. Planning for image-guided radiosurgery. In *AAAI 1994 Spring Symposium Series. Application of Computer Vision in Medical Image Processing*, Stanford University, March 1994. Also in Proc. IEEE Int. Conf. Robotics and Automation.

[Tay93] R. Taylor. An overview of computer assisted surgery research. *International Symposium on Robotic Research*, October 1993.

[TLBM94] R.H. Taylor, S. Lavallee, G.C. Burdea, and R.W. Mosge. *Computer integrated surgery.* Mit-Press, 1994. To appear.

[Zha94] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *IJCV*, 13(2):119–152, 1994. Also RR No.1658, INRIA, 1992.